Distracted Driving from use of Electronic Communication or Other Electronic Devices
2016-2020 Data

David B. Brown, PhD, P.E.
brown@cs.ua.edu

Table of Contents

Introduction

The purpose of this report is to document the findings of a study that was launched in an experiment to determine if a set of Machine Learning weighting factors could be effectively generated for two Distracted Driving (DD) categories: Code 2, Distracted by Use of a Communication Device, and Code 3, Distracted Other Electronic Device. These two categories will generally be considered together and referenced as DDED, Distracted Driving by means of an Electronic Device. These devices will be referenced as eDevices.

This study served three purposes:

1. To determine an initial set of weighting factors that might be of use in subsequent Machine Learning exercised;
2. To generate, as a byproduct, those attribute values that are over-represented in DDED crashes, and thus to provide insight into DDED crash countermeasures; and
3. To provide an estimate of the number and the actual crashes that could be inferred to have been miss-coded leading to the possibility of correcting under-reporting of DDED crashes.

The first two of these purposes will be accomplished simultaneously, since the weighting factors will be determined by over-representations in several independent variables (or attributes) that have been found to have significant over-representations. These attributes will be presented in the next section, beginning with C010, the Rural/Urban indicator. That section will further present a rationale for these over-representations based on the 2016-2020 data. The Odds Ratios taken from identical IMPACT analyses will be used as the initial weighting factors for the machine learning process. However, only one year, 2019, was used in these analyses since a five-year database would be impossible to handle with Excel.

These findings are augmented in the next section with a review and update of a previous study of DD that was performed using 2012-2016, which is on SafeHomeAlabama.gov here:
http://www.safehomealabama.gov/wp-content/uploads/2018/12/Distracted-Driving-2012-6-Data-v06.pdf
All of these conclusions were checked and those still valid are included as originally stated. Those that have changed are updated to provide the most recent findings.

The section after that, entitled Inference to Correct Under-Reporting, presents the technical details involve in transforming the data as processed above into a form that will support generating these inferences. Finally, the IMPACT analyses themselves for the 2016-2020 time frame are given in the final section.

Summary of Findings from Over-Representations in Recent Analysis

This section points out the over-representations that were found by the IMPACT analyses, and it provides some rationale for these findings. All of the findings listed were over-represented as determined by a statistically significant Odds Ratio. For this reason, they were chosen as good seeds for weighting factors for an initial machine learning iteration.

C020 Distracted Driving – Test and Control Subsets

This presents the values for DD that were contained within the test and control subsets that the IMPACTs are comparing for other relevant attributes. It is presented first to provide an overall perspective on the IMPACT comparisons that were made. The only DD attributes that were of concern in this study were those within what we defined as DDED above (distractions from phones or other electronic devices). Generally, these are considered to be the most significant of the distractions as evidenced by laws that have been passed in many jurisdictions. We consider them separate from all others because the countermeasures devised for them have little effect on the other DD attributes. The findings will now be presented in terms of the crash attributes.

C010 Rural or Urban

DDED were over-represented in rural areas by a significant Odds Ratio of 1.285 above that expected when compared to all other crashes during the five year period from 2016-2020. We reason that people are more inclined to (1) use their eDevices and (2) not pay as much attention to the roadway in areas where the traffic is not as dense. Other factors below tend to confirm this finding, which indicates that selective enforcement for DD should not be limited to urban areas.

C023 Manner of Crash

Three values of Manner of Crash were significantly over-represented in the DDED analysis (Odds Ratio, Max Gain):

- Rear End Front to Rear (1.506, 2414.3) – it is reasonable to see that drivers who are distracted will have the tendency to run into others who are stopped (e.g., at a stop sign, a traffic signal, or any other reason).
- Single Vehicle Crash All Types (1.203, 532.5) – distracted drivers frequently run of the road and lose control, hitting whatever object is in their way.
- Head-On Front to Front (1.228, 64.7) – straying into the opposite lane would also be common, leading to crashes that have the highest severity of any Manner of Crash.

C025 Crash Severity

The crash severity for DDEDs was not significantly different from non-DD crashes in the fatal category. However, it was over-represented in all of the other injury severity categories, with

Suspected Minor Injury and Possible Injury both being statistically significant in their over-representations. The conclusion to be drawn is that crashes caused by DD are generally more severe than crashes in general. This occurs because the driver is not in a favorable position to take evasive action to lessen the severity. The cause of the crash is generally the cause of the increased severity, even though DDED crashes are not over-represented at higher speeds.

C121 CU Driver Condition

Three driver conditions were found to be significantly over-represented (Odds Ratio, Max Gain):

- Apparently Normal (1.128, 1387.9) – while this is probably not seen by many as being a causal factor, it may reveal the tendency of those without any negative condition to go ahead and use the eDevice. It correlates with slightly over 90% of the DDED crashes.
- Under the Influence of Alcohol/Drugs (1.467, 195.1) – on the other hand, it appears that those who have been indulging in intoxicants are also over-represented, perhaps due to their losing their best judgment impulses. We conclude that while other factors may make drivers apprehensive to use eDevices, apparently the use of intoxicants is not want of them.
- Emotional, Depresses/Angry/Disturbed (2.151, 47.1) – when such a person is on the phone, the judgment needed to drive safely is consumed by their attention to their conversation. This was the most highly over-represented factor in terms of the Odds Ratio. Fortunately, it only involved a limited number of drivers.

C129 CU Vehicle Maneuver

Movement essentially straight might be compared to no Apparently Normal above in that there could be a reluctance to use an eDevice when other maneuver conditions are present. However, most roadways do not remain straight for more than a few minutes before a curve must be negotiated. So these two in combination account for over 82% of the DDED crashes.

C204 CU Sequence of Events #1

We saw above, with Manner of Crash, that 3,161 of these crashes were single vehicle (over the 2016-2020-time period). This attribute confirms that number with the collective sum of Ran-Off-Road and Crossed Centerline.

C225 CU Vehicle Damage

Major and Disabled was the only value found to be over-represented in this attribute, indicating that DDED crashes tend to have higher severities, which is consistent with our conclusions for severity given above. Again, the reason for this is probably the inability of the distracted driver to take actions to mitigate the crash (e.g., slamming on the brakes).

C227 CU Vehicle Towed

Vehicle towed because of its being disabled is significantly over-represented, consistent with the findings immediately above.

C230 CU Areas Damaged

The three over-represented areas damaged will be useful for providing evidence that other crashes not indicated to be DDED were in fact likely to have been in this category. The areas of Head-On-Center (area 12), Totaled (area 16) and Right-Front Angle (area 1) will provide indicators that the crash was likely to have been caused by DDED.

C233 CU Point of Initial Impact

Consistent with the above is the significant over-representation in Head-On-Center point of initial impact. While it is difficult to see how "Top" could be the initial point of impact, this does indicate that the vehicle rolled over and it would correlate highly with the vehicle being totaled.

C403 CU Roadway Condition

It has been well established that wet conditions lead to fewer fatal crashes, and this has typically been believed to be due to the reduction in speed in these situations. However, the reluctance to use an electronic device could also be a major factor in fatality reduction in wet weather conditions. There was a 71.8% reduction in the proportion of crashes as a result of wet weather.

C412 CU Trafficway Lanes

While it would seem that the fewer lanes on two-lane highways would have a deterrent effect on the use of cell phones, the inability for the distracted driver (or others) to recover from erratic driving seems to overcome this, resulting Two-Lane roads being the only ones with a significant over-representation. Another factor might be the absence of a shoulder or a clear roadside to allow for recovery on rural (e.g., County) roads.

Summary of Findings from Over-Representations in the Previous Analysis

This section reviews the over-representations that were found by the previous IMPACT analyses that were based on 2012-2016 data. These were in-depth analyses and we felt it was better the update and summarize them in this report so that this information is available. Those cases where there was a change indicated by the 2016-2020 data will be noted and the reason will be provided as best it can be determined from recent studies. Otherwise, "No change" will be noted. The ordering of the results will be as they appear in the original study.

C025 Crash Severity. No change, see above. – generally DDED crashes are more severe than non-DDED crashes.

C129 CU Vehicle Maneuver. No change, see above – crashes were shown to be down in obvious un-safe situations, and up where the driver feels comfortable using the phone or other electronic device.

C224 CU Estimated Speed at Impact. Drivers who use DDED devices do not seem to visualize speed as being as hazardous as curves or other roadway issues. This is especially true in the 71-85 MPH range, which we suspect would be caused by speeding on Interstate highways.

C122 CU Driver Officer Opinion Alcohol (Compare with C121 above in the previous section). No change. It was confirmed that DDED drivers were more likely to be under the influence of either alcohol or drugs (C123).

C121 CU Driver Condition (in this case Emotionally Distressed). No change, see above. The Odds Ratio for the past study was over twice (2.7), and it was about the same in the current study (2.2), showing that many emotionally distressed individuals have no aversion to using the phone while driving.

C104 CU Left Scene. In both studies the DDED drivers had significantly fewer cases where they left the scene. Perhaps they were too busy describing the crash to the person they had on the phone.

C030 Weather. No change, see C403 Roadway Condition above. DDED occurs significantly more in clear weather than it does in rainy weather.

C107 Driver Age. Recent studies confirmed that the most likely over-represented drivers are those 17-20, and above the age of about 43, the over-representation becomes under-representation, and the older the driver is (above 43) the more under-represented.

C109 Driver Gender. Male drivers have become much more over-represented than found in the previous study. The previous Odds Ratio was 1.022; it now is closer to 1.156. Both are significant, and the second is clearly significantly higher than the previous number, which indicates the trend of a greater proportion of male drivers engaging in DDED.

C110 CU Driver Residence Distance. Little change. The large majority (76%) are Less than 25 Miles from home, which probably reflects the overall traffic mix.

C001 County. No change. Counties with the largest cities tended to be at the bottom of the list with fewer than expected DDED crashes.

C002 Rural or Urban. No change. The rural areas are over-represented, which further supports the County conclusion given above.

C033 Locale. Little change. The largest number who put their DDEDs aside are in the Shopping or Business districts. The highest recent over-representations are in Open Country, Residential and School areas (ordered by highest MaxGain first). School areas were not significant in the earlier study.

C011 Highway Classification. No change. In order of worst first (by MaxGain), County, State, and Federal. Interstate and Municipal roadways tend to have the lowest proportions of crashes.

C008 Time of Day. No change. Rush hours and all night-time (dark) hours are over-represented. These are times and conditions when DDED is particularly hazardous.

C006 Day of the Week. No change. Sunday is the most over-represented (significant), and Saturday is slightly over-represented (not significantly). An analysis of time of day by day of the week shows a very strong correlation with DUI (alcohol and/or drugs).

C023 Manner of Crash. No Change – see above. The three most over-represented are Rear End, Single Vehicle and Head-On (front to front).

C017 First Harmful Event. No change. About 25% of DDED crashes are single vehicle.

C052 Number of Vehicles. Slight change. Significantly over-represented are single vehicle (25%), 3 vehicles (7%) and 4 vehicles (1%). Two-vehicle crashes are no longer significantly over-represented, but they account for 67% of all DDED crashes.

C208 CU Model Year. Change of necessity. Recent over-representations have been in the 2010 to 2014 model years.

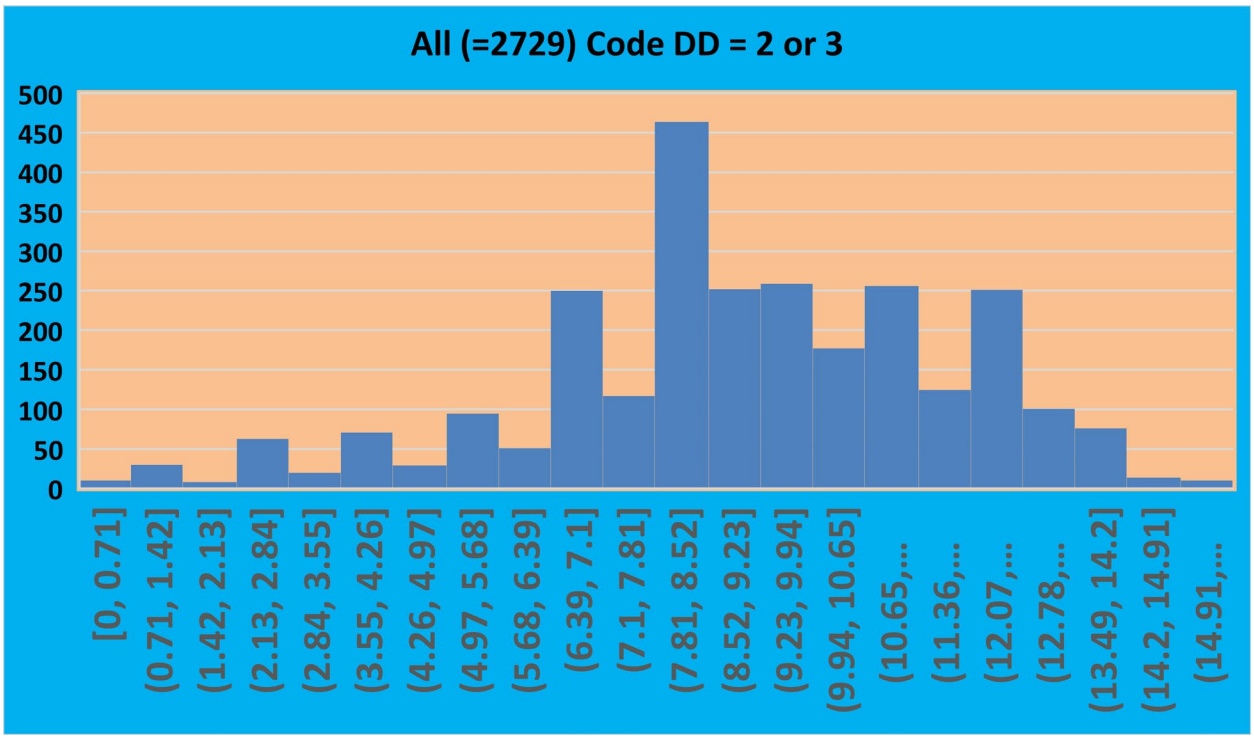
C101 Causal Unit Type. Major Change in pickups. Pickups, that were significantly under-represented, moved up to third place in over-representation, behind Passenger Cars (54%), and Sport Utility Vehicles (23%). Pickups had 18% of the DDED crashes.

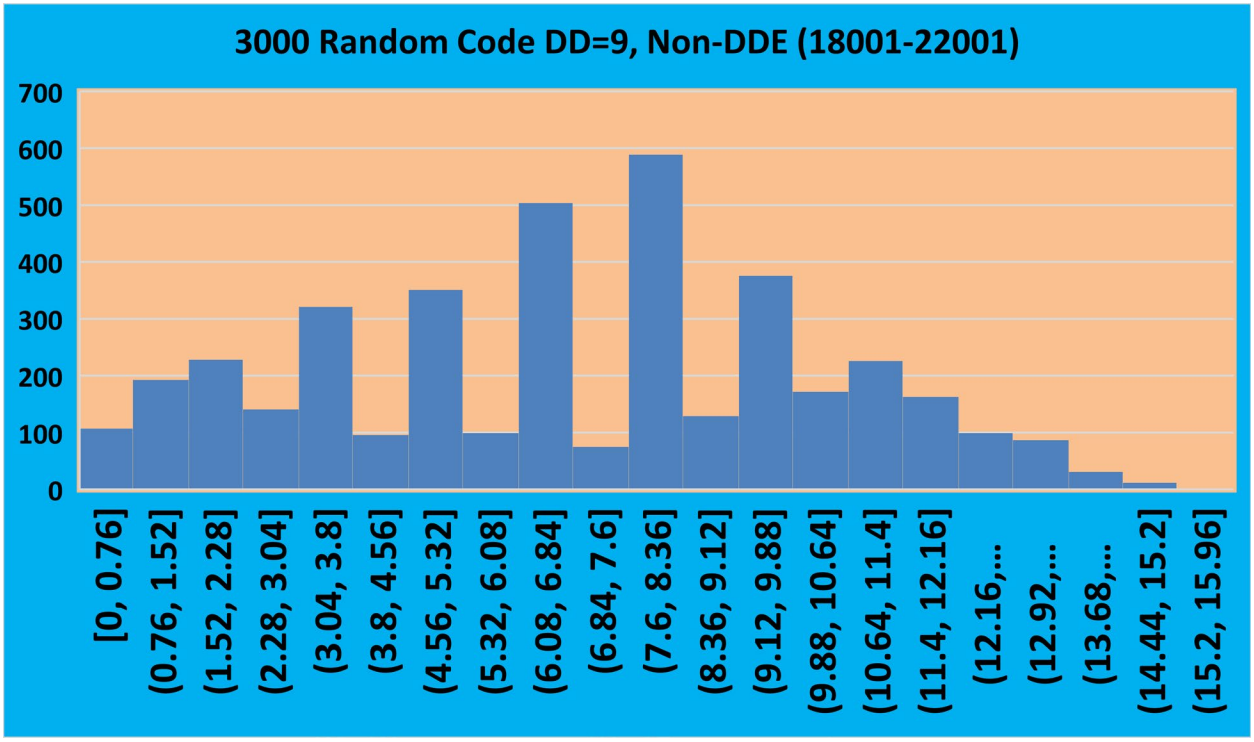
Inferences to Correct Under-Reporting

This study was set in motion by an attempt to find weighting factors to be used in subsequent machine learning with regard to DDED. It is the opinion of most traffic safety professionals that the reporting of all DD is lower than reality because of difficulty officers have in validating that DD has taken place.

The process used in creating the first step weighting factors was as follows:

1. The IMPACTs summarized above and detailed below were performed in order to determine the attribute values that were over-represented in the DDED cases. Note that the IMPACTS below were run for the entire most recent five-year period (2016-2020), while the Odds Ratios used for the initial weights were obtained from the same analyses but limited to one year, 2019.
2. A DataGen was performed on all of the 2019 data for the attributes given above. The calendar year 2019 was chosen since it appeared to contain typical and timely results. The integer option for DataGen was chosen so that the integer that relates to the specific code was captured as opposed to the value label descriptors. This greatly facilitated the substitution of the weighting factors.
3. The Odds Ratios seemed to be proportional to the over-representation for each of the attribute values that were over-represented. Since each of these was represented by a unique integer within its respective attribute, it was fairly simple to substitute the Odds Ratio that represented the degree of over-representation for the integer that was created by DataGen. We are referring to these numbers as "weighting factors" since their values are generally proportional to the over-representations that they represent.
4. In order to get an indication of how any given crash corresponded to a DDED case, the weighting factors were summed across the attributes for each of the crashes. This resulted in a total indicator for each crash. The higher this sum of weighting factors for a crash, the higher the crash was correlated to the over-represented DDED attributes. An example of these weighting factors for a number of crashes can be seen within the Excel sheet that will accompany this report.
5. The two charts below compare these totals as indicated:
    - The first chart is for DDED crashes, depicting all 2,729 DDED crashes (2 = Distracted by Use of Electronic Communication; and 3 = Distracted by the Use of Another Electronic Device).
    - The second chart is for Non-DD crashes of any type, i.e., those coded as 9 (No Distractions) by the recording officer. A sample of 3000 cases was selected at random from the total of 83,580 such (Code 9) cases in order to make the two charts comparable in number.
6. The differences between the two charts is obvious. They both have modes between 7.5 and 8.5. However, the DDED (first) chart clearly has more cases above the mode than does the random sample, and the opposite is true of the Code 9 chart.


All (=2729) Code DD = 2 or 3
500
450
400
350
300
250
200
150
100
50
0
[0, 0.71]
(0.71, 1.42]
(1.42, 2.13]
(2.13, 2.84]
(2.84, 3.55]
(3.55, 4.26]
(4.26, 4.97]
(4.97, 5.68]
(5.68, 6.39]
(6.39, 7.1]
(7.1, 7.81]
(7.81, 8.52]
(8.52, 9.23]
(9.23, 9.94]
(9.94, 10.65]
(10.65,...
(11.36,...
(12.07,...
(12.78,...
(13.49, 14.2]
(14.2, 14.91]
(14.91,...


3000 Random Code DD=9, Non-DDE (18001-22001)
700
600
500
400
300
200
100
0
[0, 0.76]
(0.76, 1.52]
(1.52, 2.28]
(2.28, 3.04]
(3.04, 3.8]
(3.8, 4.56]
(4.56, 5.32]
(5.32, 6.08]
(6.08, 6.84]
(6.84, 7.6]
(7.6, 8.36]
(8.36, 9.12]
(9.12, 9.88]
(9.88, 10.64]
(10.64, 11.4]
(11.4, 12.16]
(12.16,...
(12.92,...
(13.68,...
(14.44, 15.2]
(15.2, 15.96]

It is possible to infer that some of the cases that were coded as 9 (Not Distracted Driving) might have been coded as a 2 or 3 if the officer had more complete information. This can be done by recognizing that the crashes with the highest combined (total) weighting factors are those that are most likely to be DDED. This is because the DDED attributes had the highest over-representations in those factors most correlated to DDED.

The attached Excel spreadsheet page presents the results of a sort on the sum of the weighting factors with the highest at the top. Each line is a crash. The totals for each crash ranged from 0 to 16.109. Those retained in this spreadsheet are only those with totals above 15.000. This is a small part of the sort of all 150,269 cases for 2019 (all crashes)

Of interest here are those crashes for which the DD attribute was marked to be 9 (Not Applicable – Not Distracted). The following is a high-level explanation for the columns in this spreadsheet:

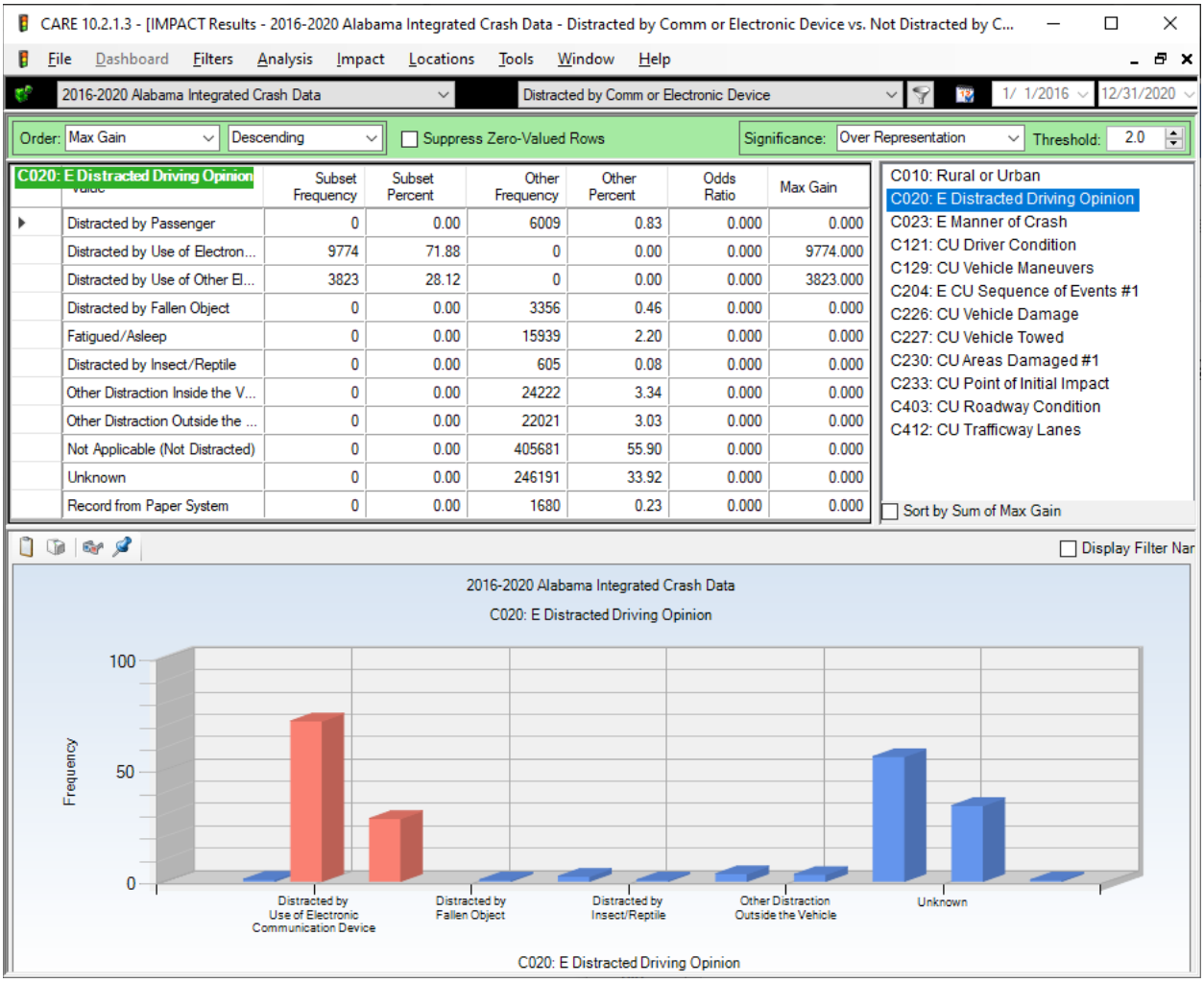
A – the crash identifier number (not used in this study).
B – Value assigned to C020 (Distracted Driving). Potential values are to the right of the listing.
All DD crashes (i.e., B column = 2 or 3 were removed prior to sorting.
C-N – Integer value of each respective attribute replaced by its Odd's Ratio = weighting factor.
C-N corresponds to Rural/Urban, Manner of Crash, etc. as given below and in Findings.
N&O – the sum of all of the weights for the crash; N and O are identical columns.

Thus, what we have is the actual code for DD that was assigned to these crashes. These may include DD codes other than 2 or 3. The interpretation of each is given to the right of the listing. The 9 values are of most interest because they are the ones that the reporting officer marked as not being any type of DD.

The 9's are marked with a yellow background, and 85 of them were found in this group of 129 crashes that had the highest sum of weighting factors (all 15 or greater). While this might seem like a large number having the highest correlations with DDED crashes that we so marked, it amounts to only 3.11% of those for which a 2 or 3 was assigned. We believe that this is a fairly conservative number of false negatives in that none of the lower-summed crashes were even considered. While some of the 9's in the 15+ category might well not be false negatives, for sure some of those in the 15- category would also be suspect. We feel that confining the estimate to just those crashes that exhibited all of the DDED over-representations is a good compromise that accounts for all possibilities.

IMPACT Displays

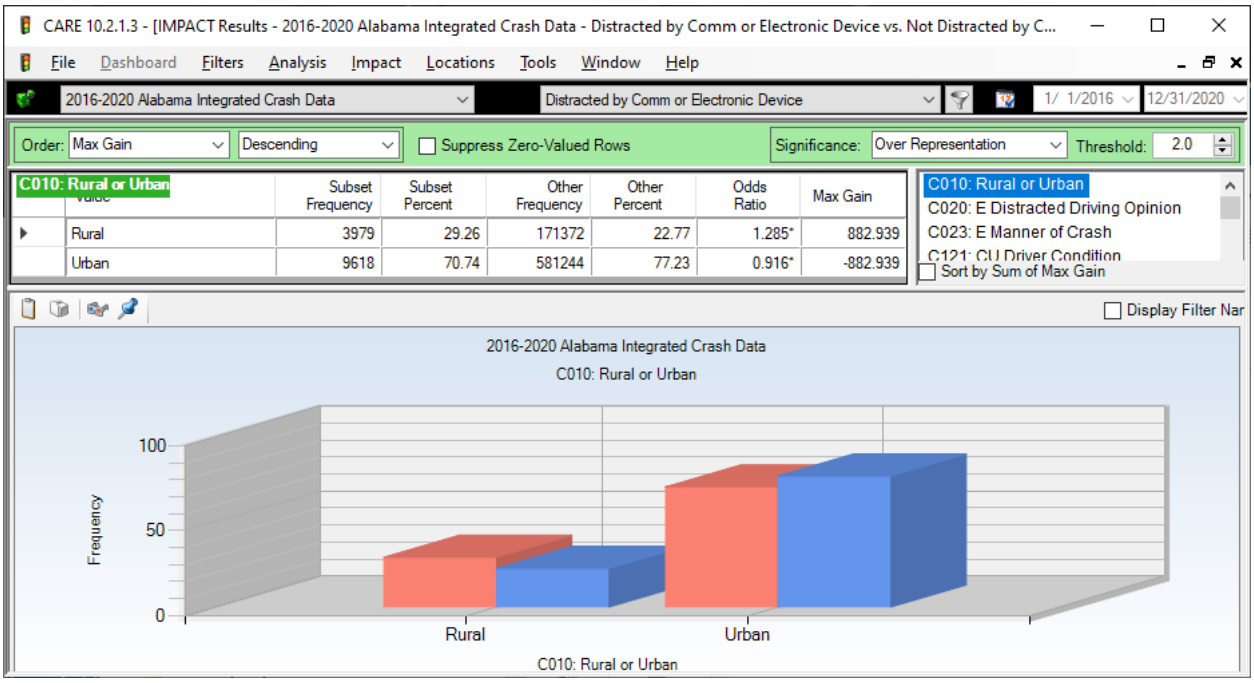
C020 Distracted Driving – Test (called Subset) and Control Subset (called Other)



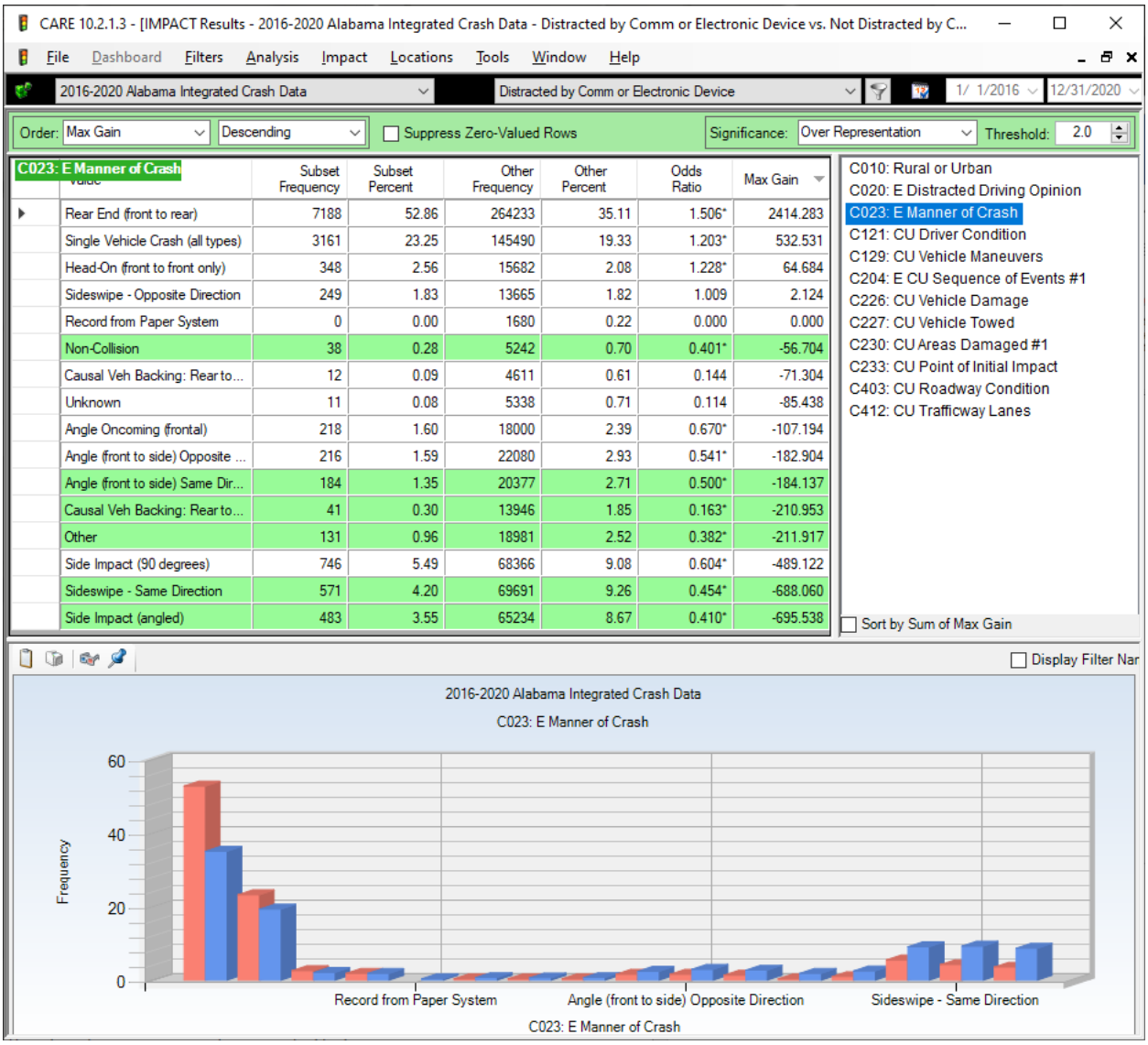
This chart is for all crashes in 2016-2020. It is given to demonstrate the data on DDED over this five-year period. It is important to realize that the IMPACTs that were summarized in the spreadsheets and used for (1) the Summary of Findings and (2) the Inference Regarding Under Reporting were obtained strictly from the 2019 calendar year.

All of the IMPACT analyses below were from the 2016-2020 data. It was felt that these five years of data would provide more accurate results than just one year.

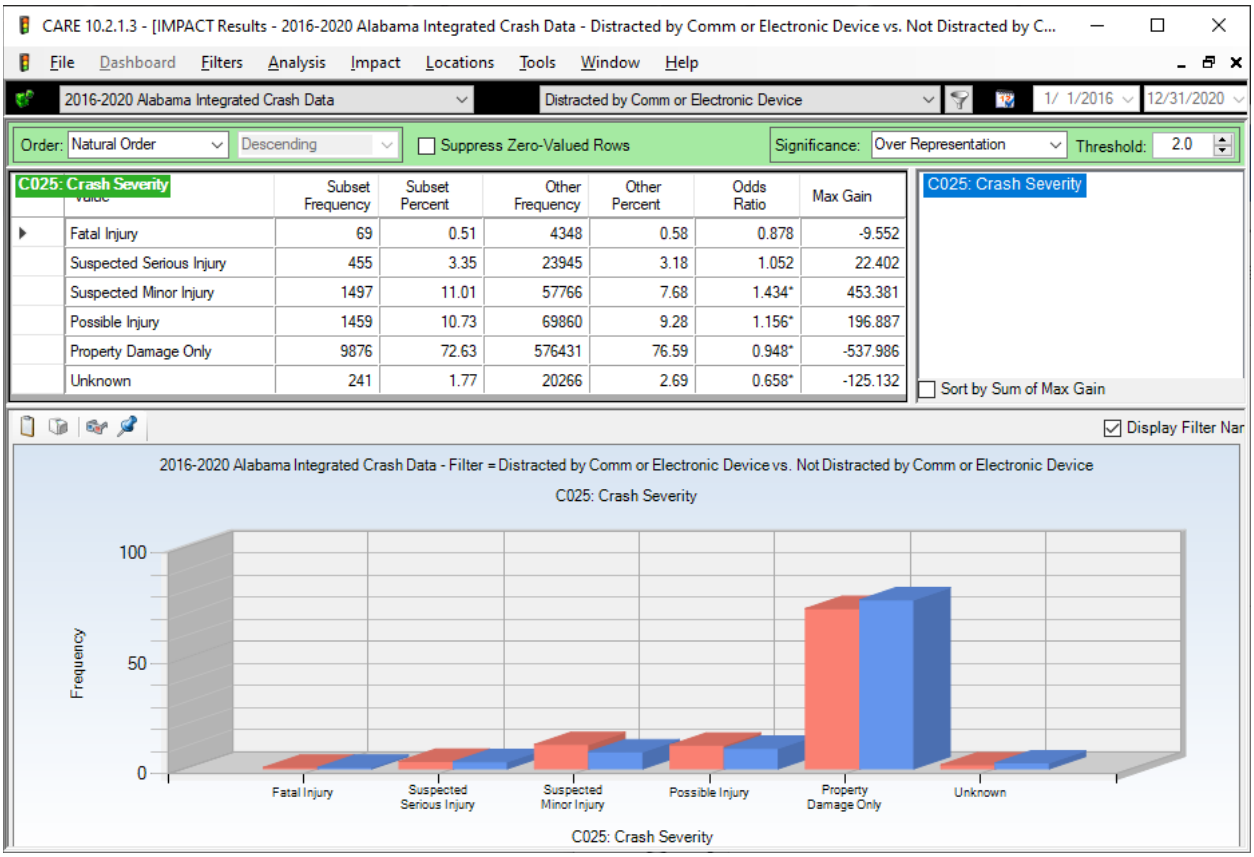
C010 Rural or Urban



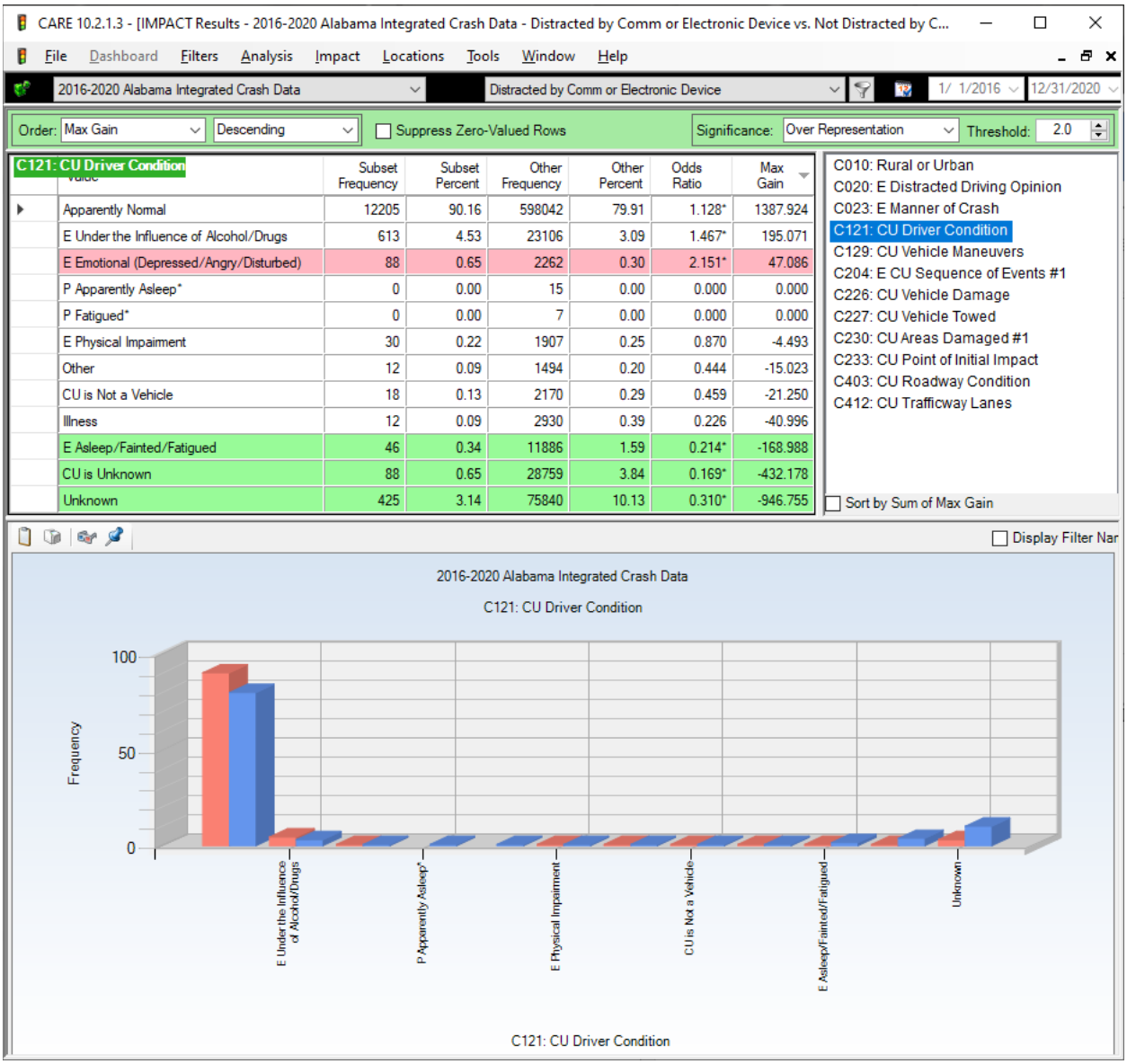
C023 Manner of Crash



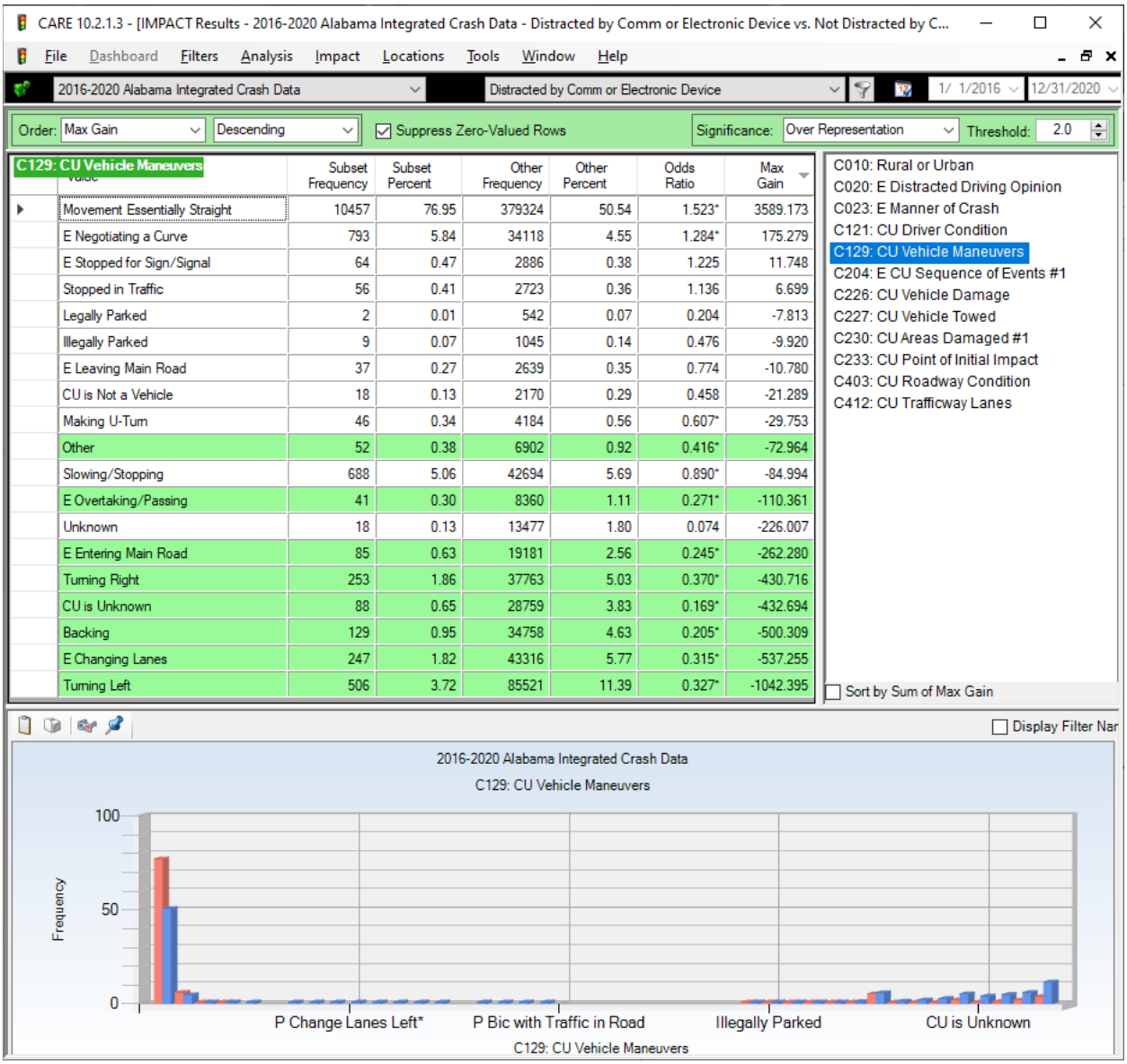
C025 Crash Severity



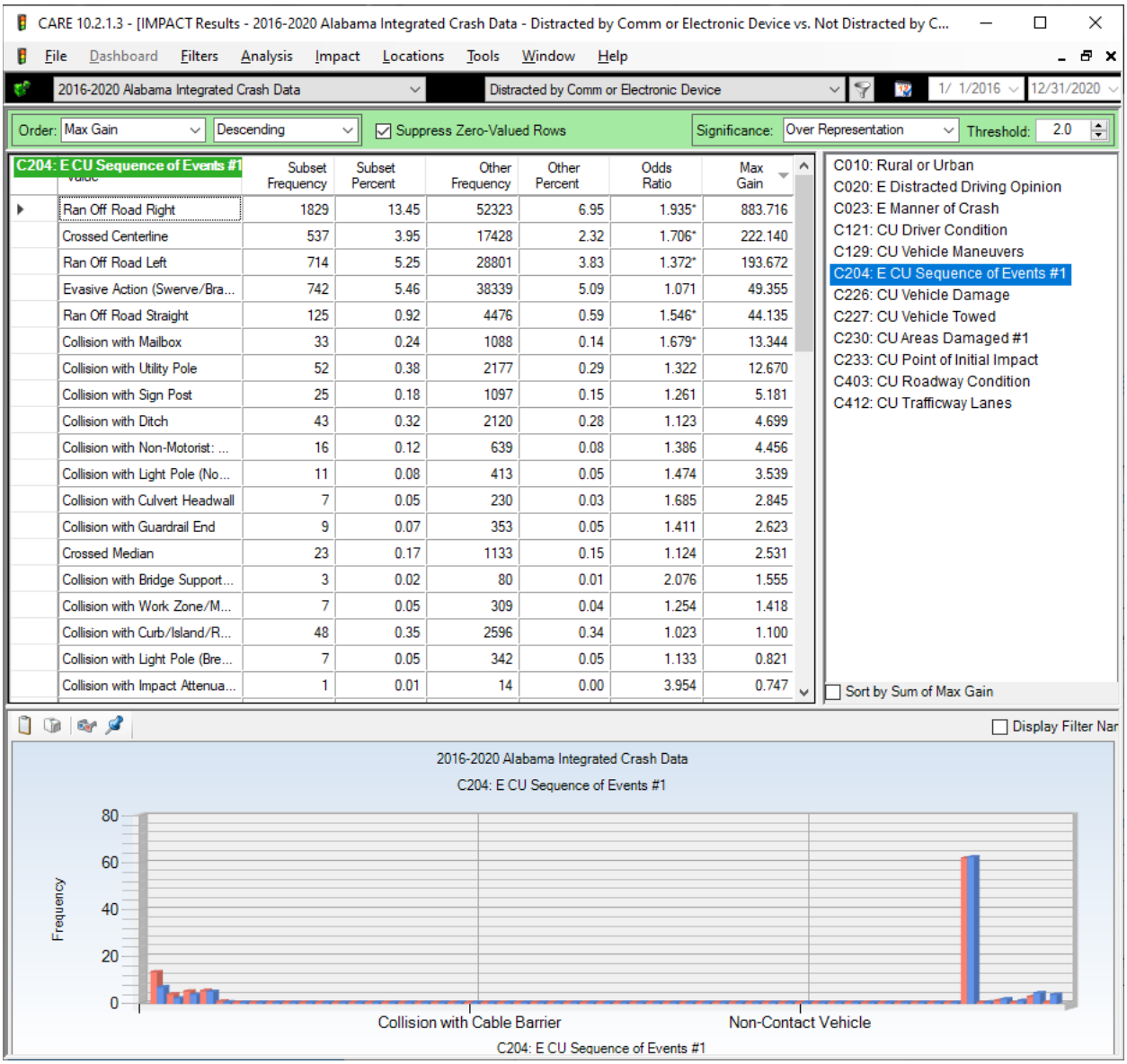
C121 CU Driver Condition



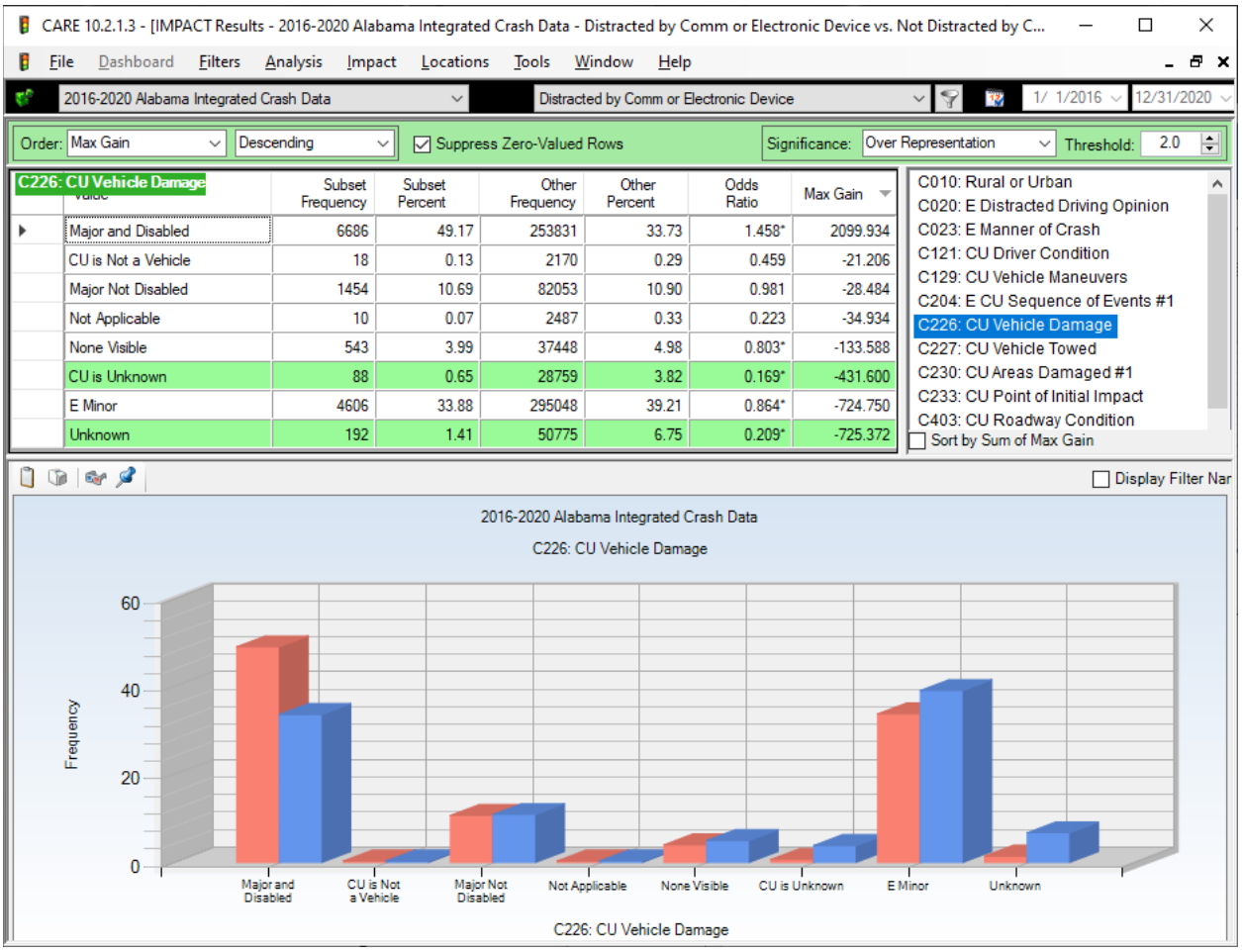
C129 CU Vehicle Maneuver



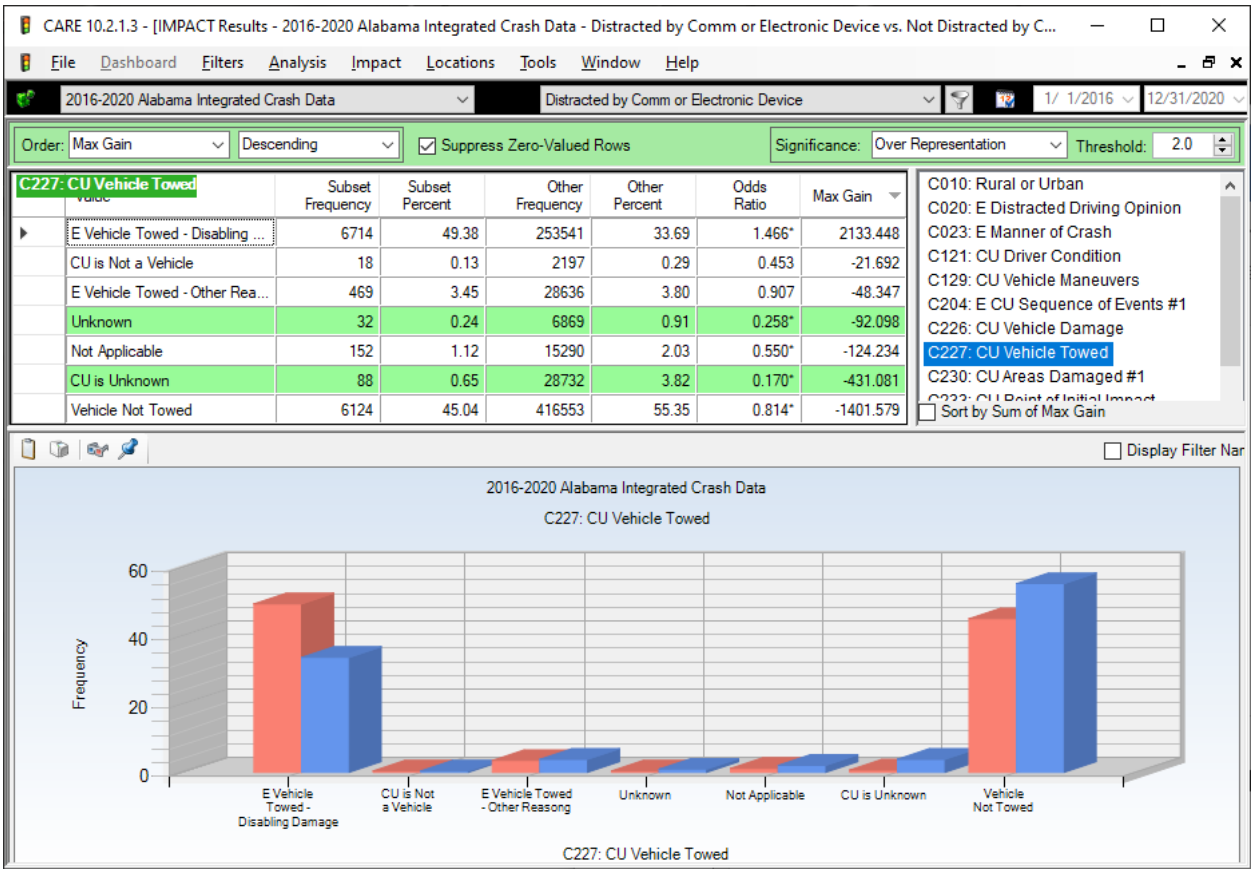
C204 CU Sequence of Events #1



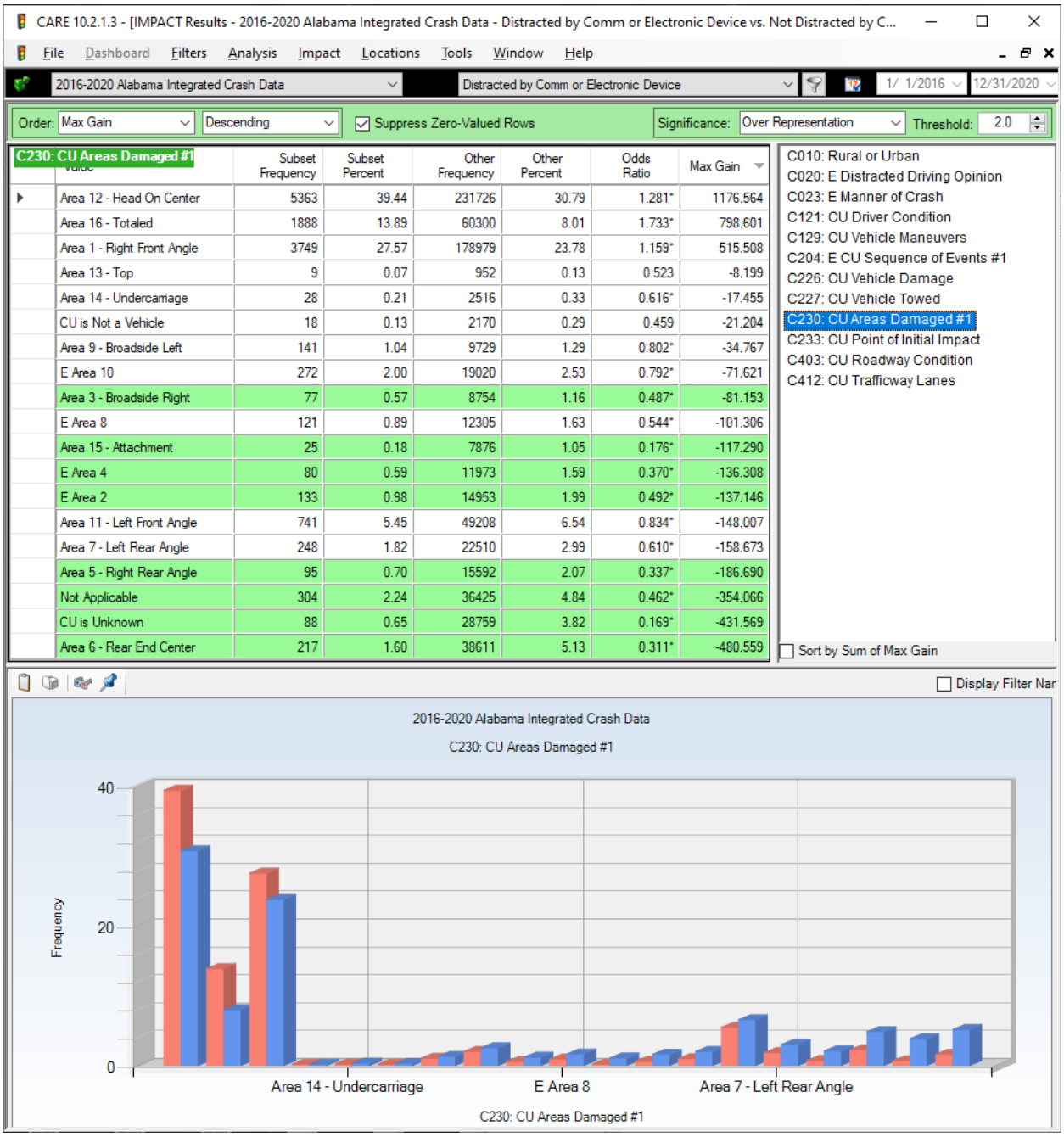
C225 CU Vehicle Damage



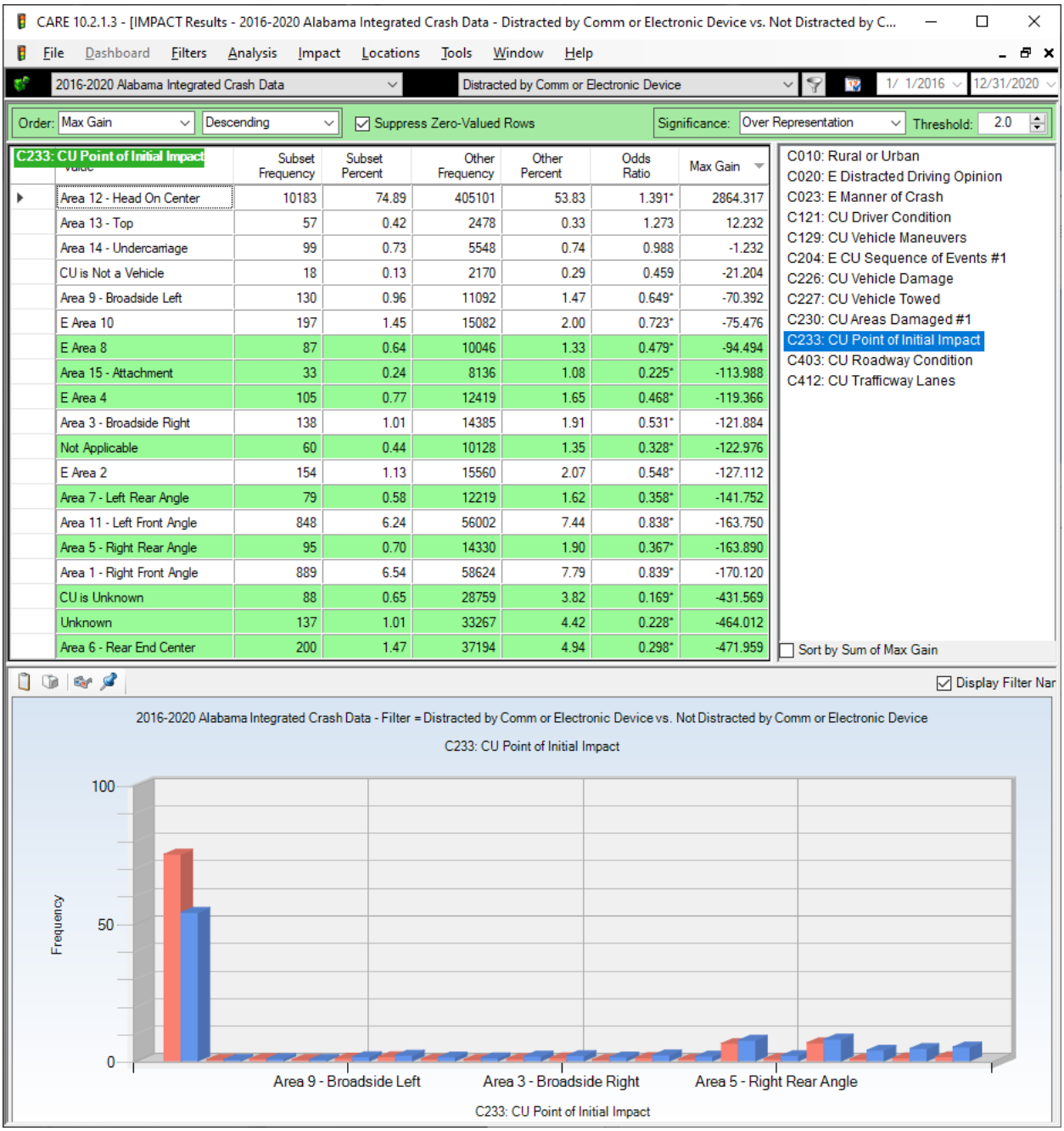
C227 CU Vehicle Towed



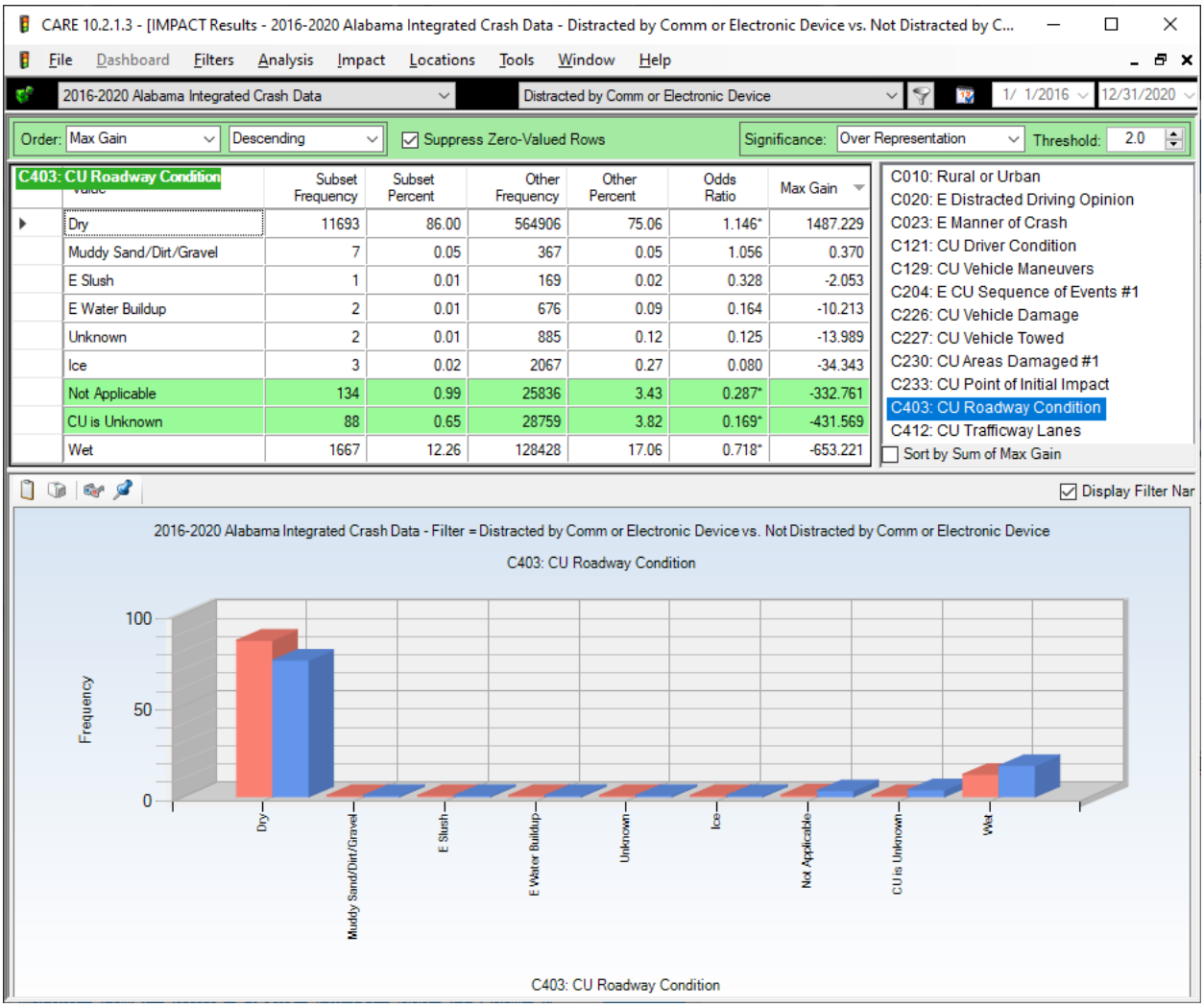
C230 CU Areas Damaged



C233 CU Point of Initial Impact



C403 CU Roadway Condition



C412 CU Trafficway Lanes